MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2019/2020

DSP5014 – STATISTICS & PROBABILITY

(RS)

15 OCTOBER 2019
2.30 p.m – 4.30 p.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This Question paper consists of 4 pages excluding cover page and appendix with 4 Questions only.

2. Attempt **ALL** questions. The distribution of the marks for each question is given.

3. Please write all your answer in the Answer Booklet provided.

4. Key formulae are given in the Appendix.

QUESTION 1

a) For the following data, state whether it is **Quantitative (T)** or **Qualitative (L)**.

No	Data
i)	The weight of a newborn baby.
ii)	Brands of television found in market.
iii)	The number of patients over 65 years old.
iv)	Types of doughnut flavours in a cafe.
v)	The number of criminal cases in a month.

(5 marks)

b) The following is the sample of 10 quiz marks:

45, 48, 52, 54, 55, 55, 57, 59, 60, 65

i) Construct a stem-and-leaf diagram for the above data. (3 marks)

ii) Find the Q_1, Q_2 and Q_3 (quartiles). (4 marks)

iii) Find the IQR (interquartile range). (2 marks)

iv) Find the lower and upper limit. (4 marks)

v) Sketch a boxplot to represent the data above. (3 marks)

c) Given the test marks for 100 students of 2 different classes as follows:

Class A

Marks	$20 \leqslant 30$	$30 \leqslant 40$	$40 \leqslant 50$	$50 \leqslant 60$	$60 \leqslant 70$	$70 \leqslant 80$	$80 \leqslant 90$
Frequency	4	10	20	30	25	10	1

Class B

Marks	$20 \leqslant 30$	$30 \leqslant 40$	$40 \leqslant 50$	$50 \leqslant 60$	$60 \leqslant 70$	$70 \leqslant 80$	$80 \leqslant 90$
Frequency	10	20	30	20	15	4	1

i) Find the mean for **EACH** class. (6 marks)

ii) Find the standard deviation for **EACH** class. (6 marks)

iii) Based on your answer in (i), which class is better? (2 marks)

[TOTAL: 35 MARKS]

Continued ...

QUESTION 2

a) Given that $P(R) = 0.75$, $P(Q) = 0.38$ and $P(R \cap Q) = 0.25$. Find

i) $P(\overline{R})$. (2 marks)

ii) $P(R \cup Q)$. (3 marks)

b) Given in a box there are two types of oranges with Grade A & Grade B as follows:

- Grade A contains 10 oranges including 4 spoil oranges.
- Grade B contains 6 oranges including 1 spoil orange.

i) By using S (spoil) and G (good), construct a tree diagram representing the information above. (6 marks)

ii) If an orange is randomly taken out of the box, what is the probability that the taken orange is a spoil one? (4 marks)

c) A classification of the equipment used in a manufacturing plant is given in the following table.

	EQUIPMENT USE			
WORKING STATUS	Low	Moderate	High	TOTAL
In progress	13	17	5	35
Under Repair	2	5	8	15
TOTAL	15	22	13	50

Find the probability of

i) working status in progress and moderate equipment used. (2 marks)

ii) high equipment used given that the working status is under repair. (3 marks)

[TOTAL: 20 MARKS]

Continued ...

QUESTION 3

a) Lifespan time for fluorescent lights have a mean of 1600 hours with a standard deviation of 150 hours. A researcher wants to test the claim that the mean lifespan time for fluorescent lights is different. A random sample of 100 fluorescent lights has mean lifespan 1570 hours.

i) State the null and alternative hypothesis. (2 marks)

ii) Calculate the Z-test value. (2 marks)

iii) What is the P-value? (3 marks)

iv) At 5% significance level, should the null hypothesis be rejected or not rejected? (2 marks)

v) Is there enough evidence to support the claim? (1 mark)

b) It was found in a survey done by Bell that 40% of its customers are satisfied with its counter for customer service. However, a new executive wants to test this claim and select a random sample of 100 customers and found that only 37% were satisfied with the customer services. At 1% significance level, does the sample information give enough evidence that the percentage is lesser than the claim?

(10 marks)

[TOTAL: 20 MARKS]

Continued ...

QUESTION 4

The data regarding the production of wheat in tons (x) and the price of the kilo of flour (y) in Factory Roskey are given in the following table.

Wheat production (x)	Flour Price (y)
30	25
28	30
32	27
25	40
25	42
22	40
24	50
35	45
25	30
40	25

a) Calculate S_{xx}, S_{yy} and S_{xy} based on the information given. (9 marks)

b) Calculate the correlation coefficient, r and interpret the answer. (4 marks)

c) Find the linear regression equation $y = B_0 + B_1 x$. (5 marks)

d) Predict the price of the flour if the number of wheat production is 56. (2 marks)

e) Compute the value of SST, SSR and SSE. (3 marks)

f) Calculate the standard error of the estimate, S_e. (2 marks)

[TOTAL: 25 MARKS]

End of Paper

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint, f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{(\sum f_i)-1}\left[\sum m_i^2 f_i - \frac{(\sum m_i f_i)^2}{\sum f_i}\right]$ where m : class midpoint, f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$Z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where $\hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$

Continued...

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$

- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{\left(\sum x\right)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{\left(\sum y\right)^2}{n}$	$S_{xy} = \sum xy - \dfrac{\left(\sum x\right)\left(\sum y\right)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{\left(S_{xy}\right)^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error, $s_e = \sqrt{\dfrac{SSE}{n-2}}$	Regression t-Test $t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$

End of Paper